**Case Study - Promotion Effectiveness of Coupon Program**

**Use CRISP-DM model to find out Promotion Effectiveness of promoting coupons**

**1 Business Understanding**

1.1 Background

ABC company is providing their truck drivers with coupons as one of the employee incentive programs. They are analyzing the demographics and utility of this program, based on usage of different categories of coupons. The objective of this analysis is to determine the factors that effects use of coupons. Based on the acceptance rate, new categories of coupons can be included, increasing the effectiveness of the program or discontinue all or certain category of coupons.

Summary of business problem:

* Do not have correct metrics to analyze the fact that coupons are used efficiently across all the categories.
* Unable to identify factors influencing usage of coupons.

1.2 Business Goals and KPI

The business goal is to determine the effectiveness of this program provided by ABC company and to make strategic decision to continue or discontinue this program

* Provide valuable insights into the effectiveness of the program in driving desired

Behaviors among drivers. By understanding the correlation between coupon usage and customer behavior

* Enhance employee engagement, motivation, and overall job satisfaction.
* ABC Company will be better equipped to tailor future promotional efforts and incentives, ultimately fostering a more rewarding and productive work environment for its truck drivers

Top of Form

1.2 Data Mining Goals and KPI

Build and analyze the data sets and build KPI’s that aligns with business goals.

* Identify patterns and trends in coupon usage data
* Determine the factors influencing coupon acceptance.
* Provide actionable insights to improve the effectiveness of the coupon program2 Data Understanding

We have analyzed the data set using coupon.csv file and made sure the dataset contains data and metadata required for this analysis and achieve business goals

**2 Data Understanding**

‘Coupons.csv’ data file consists of 12684 rows and 26 columns. Data is collected in the form of .csv file and is in the tabular format. I have used pandas library and DataFrame named **df**. Then, the **.info()** method is called on the DataFrame to display a summary of its structure and components, including the number of rows and columns, column names, non-null counts, and data types of each column.

2.1Gathering and Describing Data

The data consists of various attributes, like destination, coupon, temperature, age, income etc and flag that shows if the coupons are accepted based on these attributes.

Following python libraries are used :

##use necessary libraries in the code

import matplotlib.pyplot as plt

import seaborn as sns

import pandas as pd

import numpy as np

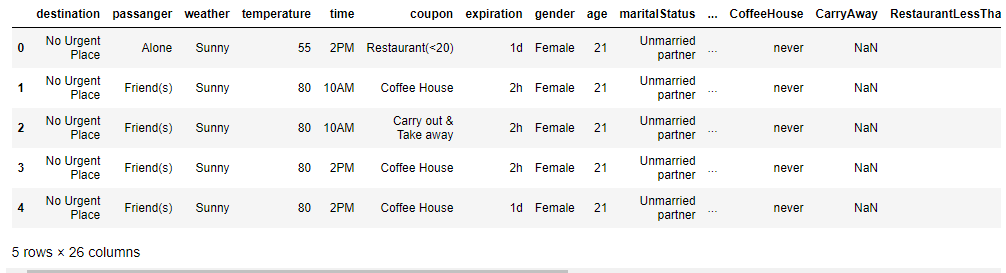
import plotly.express as px

##1. Read in the `coupons.csv` file

df = pd.read\_csv('data/coupons.csv')

##Summary of dataframe that shows the structure and components of the dataframe

df.info()



* The number of rows (each row represent a single customer data) : 12684
* The number of column : 26
* The name of each column
* The data type of each column

df.info()

**Data columns (total 26 columns):**

**# Column Non-Null Count Dtype**

--- ------ -------------- -----

0 destination 12684 non-null object

1 passanger 12684 non-null object

2 weather 12684 non-null object

3 temperature 12684 non-null int64

4 time 12684 non-null object

5 coupon 12684 non-null object

6 expiration 12684 non-null object

7 gender 12684 non-null object

8 age 12684 non-null object

9 maritalStatus 12684 non-null object

10 has\_children 12684 non-null int64

11 education 12684 non-null object

12 occupation 12684 non-null object

13 income 12684 non-null object

14 car 108 non-null object

15 Bar 12577 non-null object

16 CoffeeHouse 12467 non-null object

17 CarryAway 12533 non-null object

18 RestaurantLessThan20 12554 non-null object

19 Restaurant20To50 12495 non-null object

20 toCoupon\_GEQ5min 12684 non-null int64

21 toCoupon\_GEQ15min 12684 non-null int64

22 toCoupon\_GEQ25min 12684 non-null int64

23 direction\_same 12684 non-null int64

24 direction\_opp 12684 non-null int64

25 Y 12684 non-null int64

2.2 Data Preparation and Data Cleansing

# Get unique values from each column

unique\_values = {col: df[col].unique() for col in df.columns}

print(unique\_values)

destination : 'No Urgent Place', 'Home', 'Work'

passanger : 'Alone', 'Friend(s)', 'Kid(s)', 'Partner'

weather : 'Sunny', 'Rainy', 'Snowy'

temperature : 55, 80, 30

time : '2PM', '10AM', '6PM', '7AM', '10PM'

coupon : 'Restaurant(<20)', 'Coffee House', 'Carry out & Take away',

'Bar','Restaurant(20-50)'

Expiration : '1d', '2h'

Gender : 'Female', 'Male'

Age : '21', '46', '26', '31', '41', '50plus', '36', 'below21'

maritalStatus: 'Unmarried partner', 'Single', 'Married partner',

'Divorced', 'Widowed'

has\_children : 1, 0

…etc

##2. Investigate the dataset for missing or problematic data

##print(null\_counts) from dataset

total\_counts = (df.isna().sum()).sum()

print(total\_counts)

13219

## Identify nulls

df[df['CarryAway'].isnull()].coupon.value\_counts()

null\_counts = df.isnull().sum()

total\_percentage\_null = (null\_counts / df.size) \* 100

total\_percentage\_null = total\_percentage\_null[total\_percentage\_null != 0]

print(total\_percentage\_null)

car 3.813405

Bar 0.032445

CoffeeHouse 0.065801

RestaurantLessThan20 0.039420

Restaurant20To50 0.057310

dtype: float64

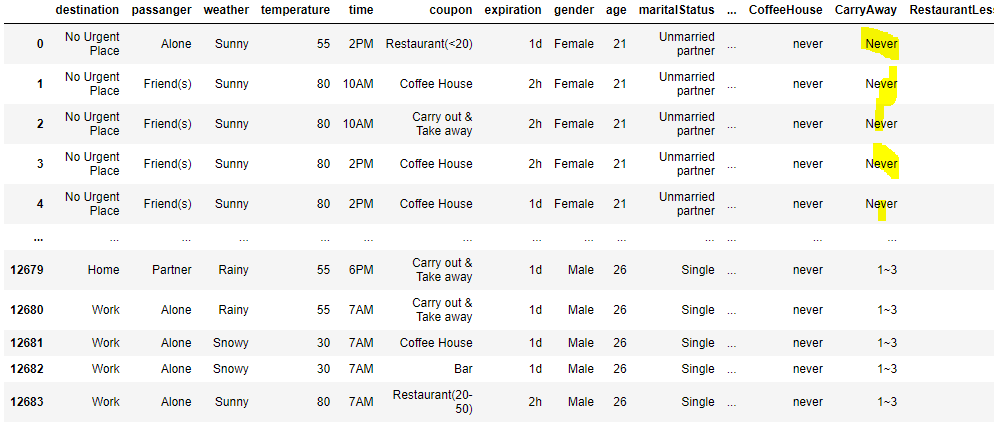
##3. Replace NaN WITH 'Never'

## Replaced NaN CarryAway values with 'Never'

df['CarryAway'].fillna('Never', inplace=True)

df['Bar'] = df['Bar'].replace('',0)

df



## Investigate the dataset for missing or problematic data

df[df['CarryAway'].isnull()].coupon.value\_counts()

Coffee House 57

Restaurant(<20) 34

Bar 24

Carry out & Take away 22

Restaurant(20-50) 14

What proportion of bar coupons were accepted? 827

if (df\_bar['Y'] == 1).any():

sum\_dfbar = df\_bar['Y'].sum()

print(sum\_dfbar)

7210

#########################################################

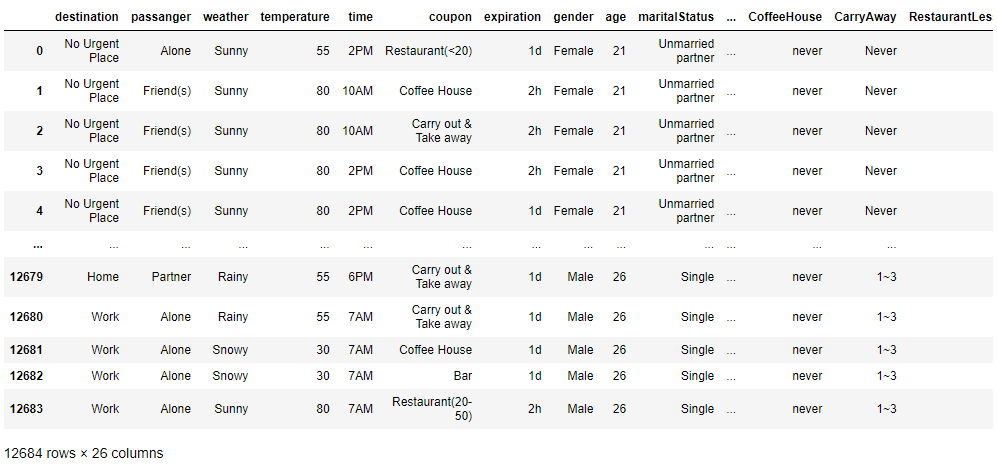
Decide what to do about your missing data –

**replaced space with ‘Never’**

## Replaced NaN CarryAway values with 'Never'

df['CarryAway'].fillna('Never', inplace=True)

df['Bar'] = df['Bar'].replace('',0)



**3. Exploratory data analysis and Visualizations**

What proportion of the total observations chose to accept the coupon?

7210 total coupons were accepted by combining all the attributes

##5. Use a bar plot to visualize the `coupon` column.

couponcounts = df['coupon'].value\_counts()

plt.figure(figsize=(12, 8))

bars = plt.bar(couponcounts.index, couponcounts.values, color='blue')

# Add values on top of the bars

for bar in bars:

yval = bar.get\_height()

plt.text(bar.get\_x() + bar.get\_width()/2, yval, int(yval), va='bottom')

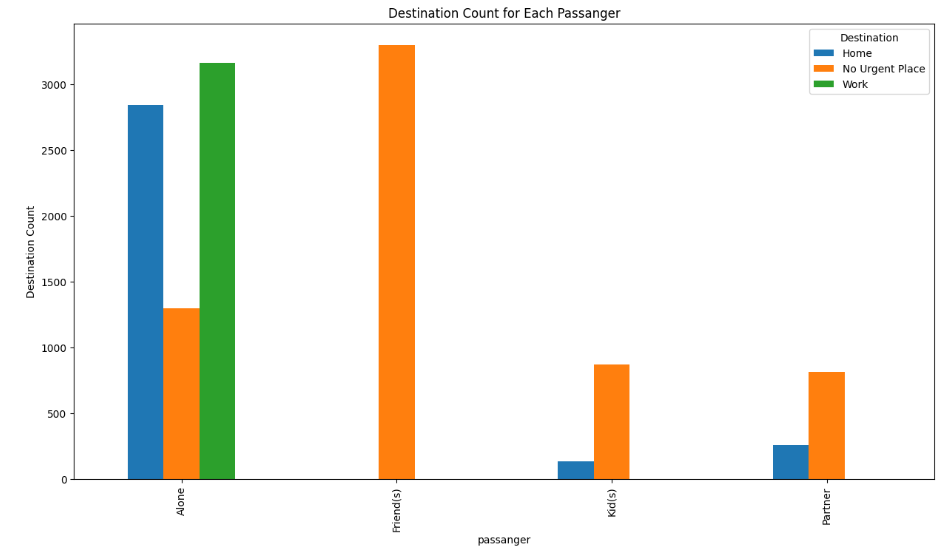
# Add labels

plt.xlabel("Coupon Category")

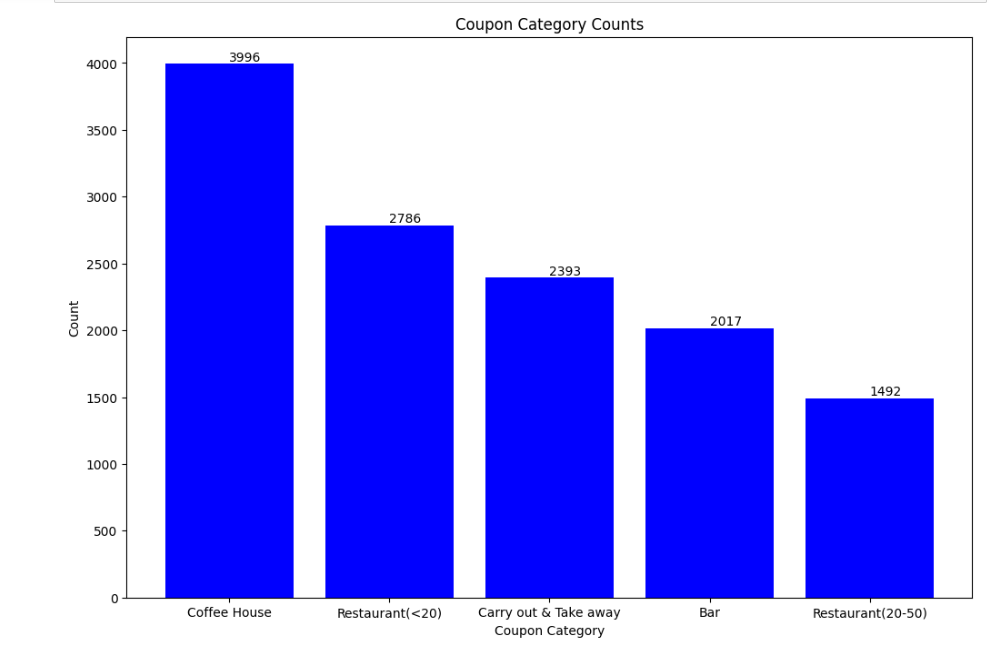
plt.ylabel("Count")

plt.title("Coupon Category Counts")

plt.show()

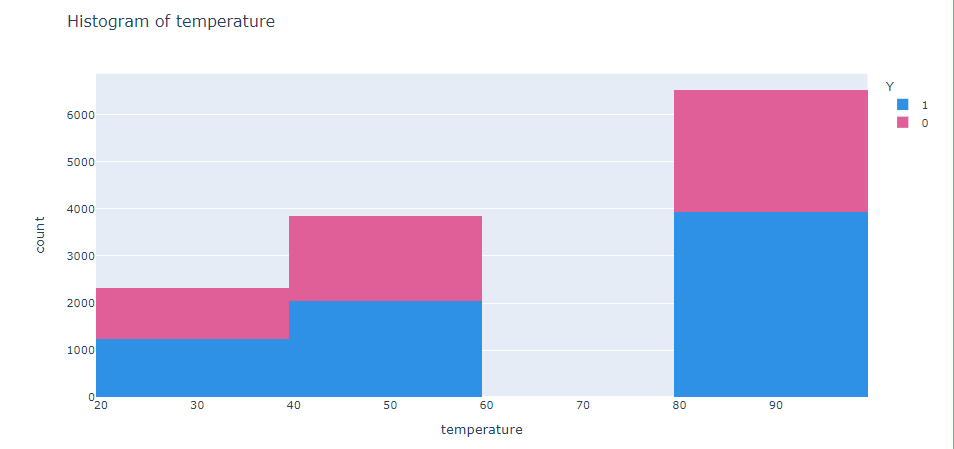


Bar Plot is used to visualize Counts by Coupon Category



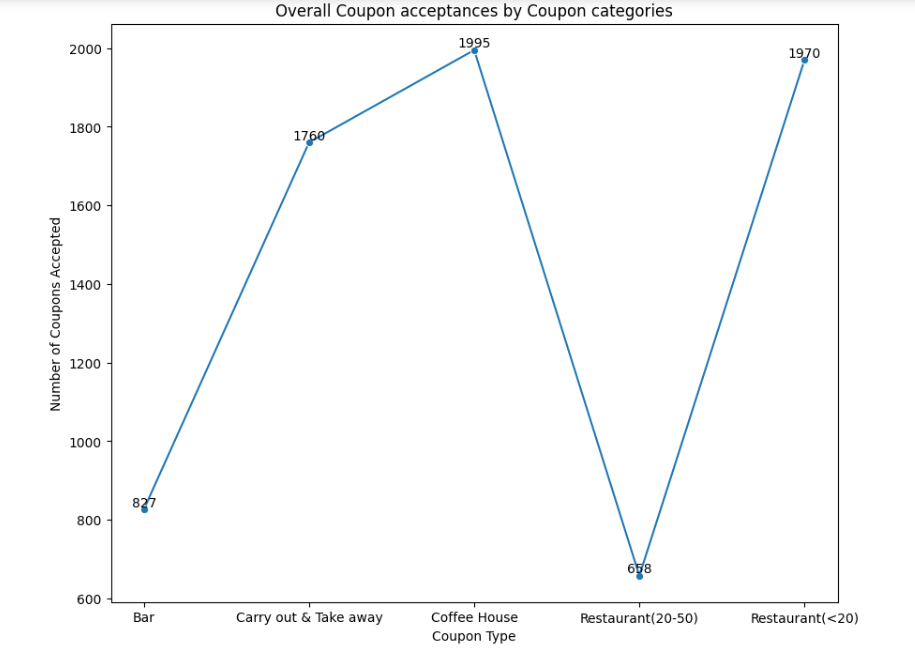
**Observation:** The coupons for ‘Coffee house’ are mostly accepted

**Use a histogram to visualize the temperature column.**



**Observation :** Coupons are accepted more when temperature is high

**Customer acceptances by overall coupon categories**



**Observation** : Less priced coupons are more accepted, Coffee House and less expensive restaurants.

Compare the acceptance rate between those who went to a bar 3 or fewer times a month to those who went more.

BarCatgroup = df\_bar.groupby('Bar')

BarCatgroup['Y'].sum()

Bar

1~3 257

4~8 117

gt8 36

less1 253

never 156

# Selecting rows where a condition is met

##I can clean the data for 1~3, however, chose to use 'As Is'

Acceptance sum for those who went to a bar 3 or fewer times a month: 666

Acceptance Rate for those who went to a bar 3 or fewer times a month: 9.237170596393897

Acceptance Rate for those who went to a bar 3 or fewer times a month rounded to 2 decimals: **9.24**

Acceptance rate between drivers who go to bars more than once a month

##and had passengers that were not a kid and had occupations other than

farming, fishing, or forestry source"

**Acceptance sum: 374**

**Acceptance Rate : 45.223700120918984**

**Acceptance Rate (rounded to 2 decimals) 45.23**

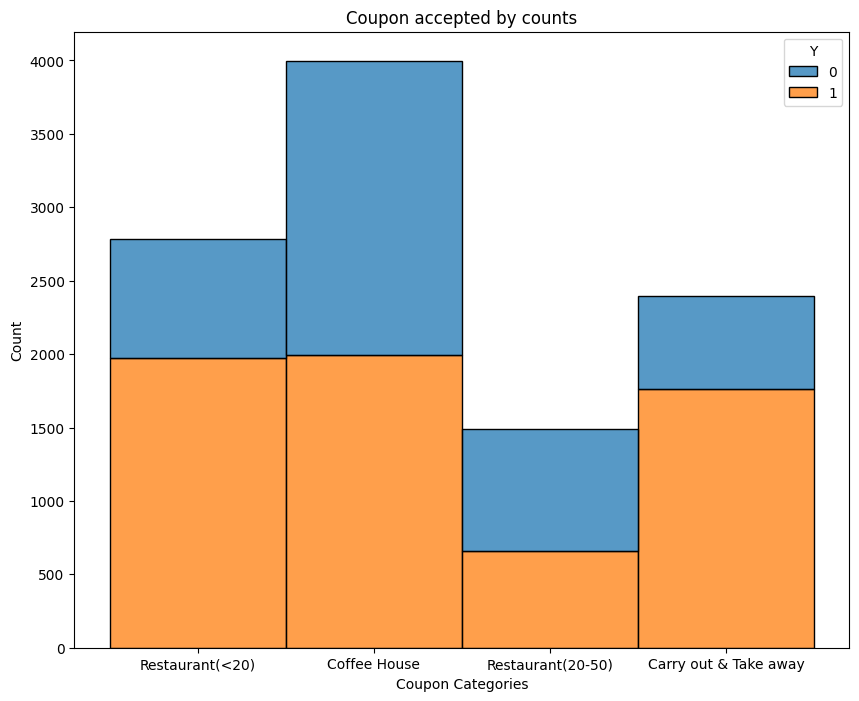
acceptance rates between those drivers who go to bars more than once a month, had passengers that were not a kid, and were not widowed and are under the age of 30 and go to cheap restaurants more than 4 times a month and income is less than 50K

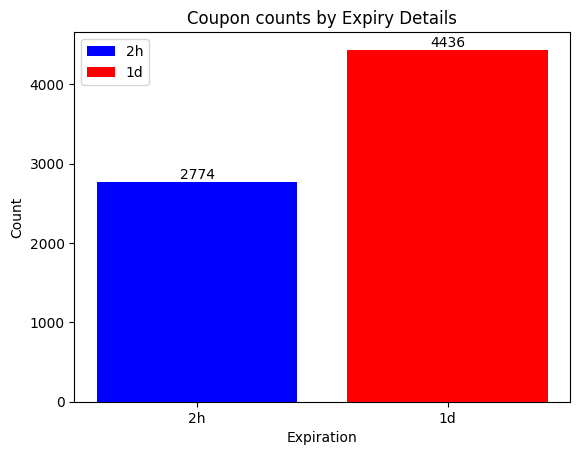
Acceptance sum[Several conditions']: 374

Acceptance Rate['Several conditions'] : 45.223700120918984

45.23

Coupon counts by Expiry Details





**Conclusion:**

Based on the observations,

* More coupons expire in a day
* Passengers go to Coffee House and Cheaper restaurants more than other categories
* More coupons are used when temperature is low
* More Females (50 Pus) buy coupons than Males
* Passengers with no kids and having company tend to buy more Bar coupons
* Passengers who are alone tend to go to work more and drive cars

**Recommendations**

Based on demographics, Coffee House and Cheaper restaurant coupons are used, based on various demographics. These coupons can be continued to be given by the company.

Coupons for expensive restaurants can be discontinued, as the usage is low

Temperature has huge influence on the coupon usage. More coupons can be sent on days when the temperatures are low and that too for Coffee house.